科学数据共享研究现状与展望

■ 毕达天[1] 曹冉[1] 杜小民[2]
[1] 吉林大学管理学院 长春 130022 [2]营口理工学院经济管理系 营口 115014

摘要：[**目的/意义**]分析科学数据共享的研究现状及其发展趋势，为促进国内相关理论研究与实践进展提供参考。[**方法/过程**] 通过对国内外文献进行梳理和归纳，提炼出与科学数据共享研究相关的三大类主题：科学数据共享发展需求、科学数据管理共享模式与机制和科学数据服务共享度与服务效果，并构建科学数据共享管理分析框架，该框架体现上述三类主题之间的因果关联，强调科学数据共享管理与服务的合理面、重要度、关键性。[**结果/结论**]国内外现有研究尚存不足，未来研究在加强实证研究、着眼管理机制完善的同时，还需注重共享服务模式的创新。

关键词：科学数据共享 数据管理 数据服务
分类号：G250
DOI：10.13266/j.issn.0252-3116.2019.24.008

科学数据是由科研活动或通过其他方式产生，反映科研活动中相关对象的本质、性质、变化规律，并支持科研活动的一类数据。随着大数据时代的来临以及大数据在各领域的广泛应用，科学发现出现了被称为“数据密集型科学”的第四种新模式，即以数据为中心和驱动、基于对海量数据的处理和分析去发现新知识为基本特征，科学数据作为国家战略资源的重要性凸显，成为支撑科技创新发展的战略资源，如海量 DNA 数据认识生命，在 1 万亿个事例中发现上帝粒子－希格斯粒子，全球碳监测网络，ARIGO 海洋浮标等[1]。为此，本研究从科学数据共享的重要作用出发，通过对当前国内外科学数据共享研究与实践的梳理、总结，分析科学数据共享的研究现状及其发展趋势，对未来我国科学数据共享提出合理展望。

1 研究设计

1.1 数据来源

数据来源为 CNKI 和 WoS 核心集，时间跨度为 2000－2019 年，检索条件为主题＝“数据共享(data sharing)”或“数据重用(data reuse)”，并含科学数据(research data)或科研数据，共检索到 73 篇文献，由文献追溯到 156 篇引文，剔除重复、不相关的篇目，最终选取高相关文献 89 篇。

1.2 分析框架构建

科学数据共享是科学数据管理的重要组成部分，科学数据共享是指通过多种形式公开发布的科学研究数据为他人所用[2]。科学数据共享源于科技创新活动和学术交流深入发展对科学数据资源的共同需求。据 A. Mulligan 等[3]对 6 344 名研究人员进行在线调查，发现 67% 的研究人员认为获取研究的支撑数据十分必要，75% 的研究人员希望获取他人的研究数据，表示愿意提供数据给他人的研究者占 52%。研究结果表明，多数研究者认为科学数据共享对科学研究具有重要意义。如何有效推动科学数据共享进程、建立健全科学数据的共享管理机制、发挥和充分挖掘科学数据的创新价值、引导和激励科研人员参与数据共享行为等问题引发了学者们的积极关注和探讨。但现有的研究分散于各个行业和各类平台共享实践中，并未得到有效的整合，在一定程度上忽略了完整因果关系的合理考量，科学数据共享研究主题没有得到有效揭示。笔者通过对国内外文献的梳理、归纳，提炼出与科学数据共享研究相关的三大类主题：科学数据共享发展需求、科学数据管理模式与机制、科学数据服务共享度与服务效果，旨在对科学数据共享过程进行整体把握和深化认识。本文构建的科学数据共享管理分析框架如图 1 所示，体现了上述三类主题间的因果关联、强调科

作者简介：毕达天(ORCID:0000－0002－9198－3799)，副教授，博士，E-mail：bdt@jlu.edu.cn；曹冉(ORCID:0000－0002－1052－0849)，硕士研究生；杜小民(ORCID:0000－0002－1741－087X)，讲师，博士。
收稿日期:2019－09－24 修回日期:2019－10－21 本文起止页码:69－77 本文责任编辑:易飞

学数据共享管理与服务的合理面、重要度、关键性。

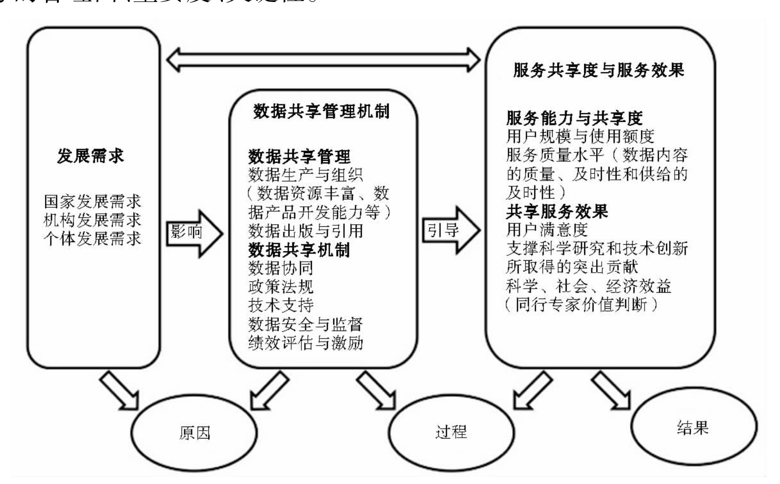


图 1　国家发展需求下科学数据共享管理框架

2　科学数据共享管理框架分析

2.1　发展需求作用下的科学数据共享

实施科学数据开放共享，是国家发展的战略需求，现已成为国际性的趋势，面向科学数据管理和共享的机构的相继成立、科学数据中心（或数据共享平台）的涌现等是这一趋势的具体体现[4-5]。

自 20 世纪 50 年代以来，随着科技界对基础科学数据的交流和应用需求的日趋显现，以及对科学数据价值认可度的提升，逐步确立了科学数据基础性地位。国际科学理事会（International Council for Science，ICSU）在 1957 年组织成立了以地球科学、空间科学和天文学数据为重点的世界数据中心（World Data Center，WDC），旨在开展地球科学、环境和空间科学领域数据的收集、整理、系统化、标准化及交流服务等活动。而后针对不同学科存在的共享障碍，国际科学理事会于 1966 年成立了覆盖更广泛学科领域的国际科技数据委员会（The Committee on Data of the International Science Council，CODATA）。2001 年 CODATA 创办了以科学数据为对象的电子杂志《数据科学杂志》（Data Science Journal），确立了科学数据成为一个新的学科领域[6]。2004 年 3 月，英国联合信息系统委员会（The Joint Information Systems Committee，JISC）启动组建了数据管理中心（Digital Curation Center，DCC），该组织积极从事科学数据的存储、管理和共享等领域的实践。2013 年 3 月，欧盟委员会（European Commission，EC）、美国国家科学基金会（National Science Foundation，NSF）和美国国家标准与技术研究院（National Institute of Standards and Technology，NIST）、澳大利亚创新部（Australian Innovation Department，AID）等共同组建了科学数据联盟（Research Data Alliance，RDA），为科研人员跨越技术和学科领域、跨越国界开放共享科学数据创造了条件，推动全球科学数据共享交换。在这些数据共享管理组织引领和推动下，各国政府逐步在国家层面上进行整体规划，建立科学数据中心体系，如美国建立了庞大的科学数据中心体系，包括 13 个学科性的世界数据中心和 9 个国家数据中心。世界数据中心包括大气微量气体中心、海洋地质学与地球物理学中心、地震学中心、人类与环境相互作用中心、气象数据中心、海洋学中心、遥感陆地数据中心、火箭与卫星中心、日地物理数据中心、地球自转中心、固体地球物理学中心以及古气候学中心等。国家数据中心典型案例是由美国国家航空航天局启动“分布式、最活跃数据档案中心群”项目，例如美国国家航天局（National Aeronautics and Space Administration，NASA）、美国国家科学基金会（National Science Foundation，NSF）、美国能源部（Department of Energy，DOE）、美国国立卫生院（National Institutes of Health，NTH）等国家级数据中心[7]。

我国首次提出科学数据共享是在 2001 年底，中国启动科学数据共享工程，涉及 24 个领域。2004 年 7 月，科技部启动国家科技基础条件平台建设，贯彻“以人为本”的理念，把国家科技创新的着力点转移到创造良好的制度环境、政策环境和科技基础条件上来。2015 年 3 月，中科曙光发布“数据中国”战略，提出“让全社会共享数据价值”的愿景理念；2015 年 8 月，国家发布《促进大数据发展行动纲要》指出要加快政府数据开放共享，推动资源整合[8]。2018 年 3 月国务院办公厅颁布的《科学数据管理办法》明确了国家科技

计划项目产生的科学数据要进行强制性汇交,体现了近年来我国越来越重视科学数据的共享[9]。时至今日,在用户基础、技术支撑、政策推进的影响下,以政府部门(及其附属机构)为主导建设的多个科学数据共享平台正式进入运行服务阶段,如国家地球系统科学数据中心、国家人口与健康科学数据中心、国家农业科学数据中心、国家地震科学数据中心、中国气象数据网(气象数据共享中心)、国家林业和草原科学数据中心[10]等。这些平台虽然在一定程度上促进了科学数据的共享与再利用,但受数据开放的时效、深度、广度等的限制,仍然难以满足快速发展的科研范式的要求。

综上,共享发展需求源于国家战略发展、科学共同体和科研团队三个层面,不同层面的共享需求能够在一定程度上激发和推动科学数据共享活动发展,最终体现了在科学数据开放共享活动中追求实现科学数据价值的共同目标。上述科学数据共享规划就是以国家发展需求为主导的整体统筹,体现了国家对实施科学数据共享工程的整体规划,它可以通过国家投入的增量而拥有整合各部门、各单位资源的能力和创建共享新秩序的能力[11]。如 20 世纪 90 年代以来,美国政府逐步在国家层面上构建数据共享的框架,将各级政府、各联邦部门、商业公司和个人的数据管理活动纳入到国家整体规划行为之中,融合形成国家级科学数据管理与服务系统。2002 年我国正式启动的科学数据共享工程,在国家的整体规划与管理下,将各部门、各单位所积累的科学数据资源纳入国家科学数据共享统一框架,旨在形成跨部门、跨学科、多层次的国家科学数据共享服务体系,实现多年积累的基础性、公益性科学数据的分类分级共享,使海量的科学数据资源的潜在价值得以充分发挥与增值。

2.2 管理情境下科学数据共享

2.2.1 科学数据共享管理模式　黄鼎成认为,科学数据共享管理系统运行的稳定性,取决于发展需求、数据资源、组织管理、共享规则、共享技术 5 个主控因素相互作用的机理,数据资源、共享规则和共享技术是共享管理的必要条件,数据资源的管理者和用户是共享平台满足发展需求而实现其效用价值的充分条件[11]。科学数据共享具有科学数据管理生命周期的过程性,与数据生产、数据组织、数据存储、数据出版、数据传播、数据引用直接关联的一种价值创造活动,体现出数据生产者、资助者、组织者、出版者、传播者、管理者、利用者等利益相关主体之间的相互协同性[12]。对于科学数据共享管理生命周期的价值创造活动而言,有两种力量促进着科学数据共享价值创造高效完成:一是科学数据共享内部动力条件,主要指数据管理机构内部诸要素彼此关联及其对外部环境变化的适应与协同程度,内部动力条件更多体现的是一个数据管理机构的数据管理能力;二是科学数据共享外部支撑条件,主要指共享活动所处的外部环境因素,主要包括共享活动中相关组织或机构(利益主体)、共享法律制度的完善程度以及信息技术环境。科学数据共享管理是学界和业界研究的重要主题之一,现有的研究大多聚焦在科学数据生产与组织,还有科学数据出版与利用等方面。

科学数据的生产与组织具有感知科学数据及其运动的认识论意义。科学数据从产生方式来看,主要来源于地面观测(监测)、遥感遥测、考察调查、统计分析,实验测试,加工处理、计算模拟,互联网挖掘等手段。在美国科学数据产出可以大致分为 3 种,分别来自政府部门、商业部门和大学与科研院所[13]。我国当前情况下科学数据的获取和积累主要是靠政府投资完成的,它的 3 个主要来源是:政府部门专项计划产生专业数据;国家各类科技计划、专项科技活动产出的专业数据;研究院所、高等院校的专业数据库以及国际科技数据组织中国分支机构所拥有的国际科学数据资源乃至科学家个人手中的科学数据等[14]。可共享的数据资源产生方式不同,为了实现真正意义上的数据共享,科研机构推动了对分散异构数据资源的整合。A. Chilvers 等[15]认为元数据是普遍意义上科学数据组织管理的关键,方便以统一的形式实现科学数据的组织和获取,成为实现科学数据与其支撑文献多维关联的有效途径。我国目前投入运行的科学数据共享平台对数据的描述采用元数据的形式[10]。科研数据组织的层次和水平是对科研数据特征与其文献关联特征认识的集中体现,科研数据中蕴含着大量的知识及关联关系,如研究主体(人员、机构、国家地区等)、科研活动(实验、项目等)、科研条件(方法、仪器、经费、期刊等)、科研产出(成果、专利、标准等)等,这些科研实体之间存在着相互交错的多维关联关系,如人员相关、机构相关、成果相关、主题相关等;同类数据资源内部也存在着继承、属分、相关等关联关系。在数据组织层面,通过合理的科学数据分类管理和系统集成,可以有力地提升科学数据的价值增值,推动数据资源的流通与共享,突出数据资源整体动态性、系统化的能动开发样式。目前,数据组织面临的主要挑战有系统异构、数据描述语法不统一、科学数据元数据格式不统一、科学

数据之间缺乏语义关联等[16]。

数据出版(data publishing)是指通过一定的公共机制发布科研数据集,使得公众根据一定规则可以发现、获取、评价和应用这些数据集[17]。科学数据出版是保证数据有效共享的重要形式之一,国内外学者针对数据出版模式[18-21]、数据出版的关键性问题[22-25]、数据出版的技术应用问题[26-28]展开了研究。科学数据出版的最终目的是数据用户能够便捷使用相关数据。B. Lawrence 等[29]指出数据出版是指数据达到可引用和可追溯的状态。因此,数据出版需要更加简单、数据可被引用、数据可方便使用[30]。

科学数据的引用是采用文献引用的方式对所使用的科学数据资源进行描述[31]。数据引用作为实现数据共享的重要途径之一,它是连接数据使用者与数据生产者的桥梁,有助于明确数据共享的责任与义务,达到对研究数据进行增值利用的目的[32]。从政策层面关注科学数据的引用是国外学术界的一个研究热点,英、美、澳、欧盟等纷纷制定了针对研究数据引用与共享的政策及实施计划。史雅莉在对国内外科学数据引用标准进行全面调研的基础上,采用案例分析法,从数据引用服务的开展、数据引用标准实施指南的联合推行、数据引用系统的开发与应用及引用标准实施效果的反馈4方面深入分析澳大利亚 ANDS 数据引用实践[33]。国外学者探讨了科学数据的引用原则[34],Dataverse 等6个出版平台认可数据引用原则联合声明(FORCE11)及"可发现(Findable)、可访问(Accessible)、可互操作(Interoperable)和可重用(Reusable)"的 FAIR 数据共享原则[35]、相关标准[36]、数据引用技术[37],如大多数的出版平台数据引用使用 DOI 和数据指纹技术(Universal Numeric Finger,UNF),并遵循国际科技数据委员会(Committee on Data for Science and Technology,CODATA)等的引用原则[38]。

国内学者对数据引用行为[39-42]、数据引用规范[43-46]、数据引用策略[47]的研究进行了系统归纳和分析,并指出我国科研人员数据引用行为在不同学科引用数据类型方面有明显差异,非原始数据的引用仍占较大比重、元数据元素的信息揭示程度不一等;由于出版商、数据仓储、科研人员等科学数据引用相关主体之间缺乏合作与交流,国内的数据引用要求引用元素的数量、种类和可拓展性都不如国外的相关要求并缺乏统一性和专业性,引用内容相对宽泛。因此,加强出版商、数据仓储、科研人员等科学数据引用相关主体之间的合作与交流,建立完整的科学数据引用体系和机制是促进科学数据引用规范的发展方向。

综上,当数据产生(观测、计算、实验方法)科学合理、数据描述信息全面(数据全面、元数据规范)、数据维护及时(数据更新状态良好)、数据获取更为方便、数据使用合乎要求(精度合理)时,用户就能感知到较高的数据价值,从而表现出较强的共享意愿。

2.2.2 数据共享机制 科学数据共享机制是指科学数据开放共享系统要素之间的相互作用过程、方式与运行规律[48]。数据共享机制对于创建数据共享的新秩序、推进数据共享平台的健康发展具有重要的作用。建立健全科学数据共享的主要机制与数据共享的5个基本要素密切相关。数据资源组织管理、共享规则的建立与共享技术的实现需要相应的数据资源协同机制、政策与法规保障机制、技术支持机制、绩效评估与激励机制、数据安全监管机制等的保障。科学数据共享的研究较早与数据共享的技术支持研究有关[49-52],随着国内外数据共享平台投入运行,数据共享规则的政策、法规问题成为研究的热点[53-56],且已取得一定的研究成果。张晓青等[57]在对国外科学数据开放共享政策进行分析的基础上,将其内容归纳为三类要素:①政策要求,包括对数据质量与标准、数据获取与共享、数据保护与保存、数据管理计划的要求;②政策规定,包括对隐私、传统知识、敏感数据、知识产权或数据所有权的规定;③其他说明,包括对科学数据开放共享目的、原则、范围、角色与责任、监管和实施的描述。从国内科学数据开放共享实践来看,我国科学数据共享服务虽已初见成效,但是由于未制定国家或行业层次的科学数据开放共享政策,影响了我国科学数据共享实践的有序发展。因此,建立一整套内容相互衔接、体系配套、效力相互补充的科学数据共享政策、法规体系将是我们今后研究工作的努力方向。

科学数据共享过程与数据生产者、资助者、组织者、出版者、传播者、管理者、利用者等利益相关主体之间是相互促进、共同依存的,通过数据来驱动相关主体之间的协同,凝练共识,聚分力为合力,应逐步建立健全包括国家层次、行业层次、机构或企业层次的科学数据开放共享政策体系,为部门间(数据采集范围的分工)协同、中央与地方的协同、科学家间的协同、数据生产者与管理者间的协同提供有效指导。此外,将数据共享机制融入科学数据共享的环境,保证数据流及时、准确、全面流通的同时,也激发、引导、强化了科研人员参与数据共享的效果。

2.3 服务共享度与服务效果

科学数据服务是一种动态服务、跨界服务和嵌入服务，既符合国家发展的需求、科研团队的需求、科研人员的需求，也符合科研支持的数据服务机构自身发展的需要[58]。数据共享服务的根本任务是连接数据拥有者和数据用户，促进数据资源的传播和重用。2011年，科技部与财政部制定了《国家科技基础条件平台运行服务绩效考核指标》，重点考察科技平台的服务数量与服务成效。该指标包括服务数量、服务成效、运行管理、资源整合4个一级指标和12个二级指标[59]。2011年4月，国家科技基础条件平台建设基础科学数据共享网项目组提出了《共享服务评价指标体系》，设计了包括数据资源建设能力、技术能力、组织管理、共享服务、效果、特色在内的6个一级指标以及20个二级指标[60]。随着我国科学数据共享工程的启动，有关数据服务共享绩效评估的研究与实践已引起了学者与业界的关注[61－62]。近年来，受到“用户满意导向”[63－64]服务绩效观的影响，数据共享服务评价从开始关注需求端，转向“以用户满意为导向”的服务效果[65－66]，并将满足用户需求的服务能力作为评价服务效果的重要标准[67]。

2.3.1 服务能力与共享度 服务能力与共享度体现了一个数据管理机构用户覆盖面的宽度、用户使用频度、提高数据的质量与时效性以及供给的及时性等服务质量。用户覆盖面的宽度、用户使用频度与数据共享服务定位有关，NSF将科学数据分为研究型数据(research data)、资源型数据(resource data)、参考型数据(reference data)3种类型。从服务定位来看，研究型数据是研究所需数据集，体现的是研究团队的价值，只能为特定的研究群体(项目研究团队)服务；资源型数据是关键的研究数据和科学共同体需要的数据集，体现的是科学共同体的价值，服务于特定的研究群体(某一领域或学科的组织和机构或跨系统组织机构)；参考型数据是对于国家和社会重要的数据集，体现的是社会价值，可以服务于整个科学界，参考型数据用户群体大、应用影响面广，开放共享要求也较资源型和研究型数据高[68－69]。从数据共享与利用行为来看，科研团队的数据共享与利用行为特征是为了科研团队在具体研究方向上的利用，数据在科研团队的协同开发利用中得以发挥科研价值，具有较强的专指性。吴丹等[70]对我国医学从业者科学数据共享行为进行了调查研究，结果表明，多数科研工作者的科研数据共享范围只愿在团队内部。科学共同体的数据共享与利用特征是以协同创新目标为导向，整合共同体成员组织的优势资源并实现共享利用，具有较强的规划性[71]。

科学数据质量是保证科学数据开放共享的先决条件。基础科学数据共享网关于服务质量的评估指标主要涉及数据内容的质量、时效性与供给的及时性3个方面。其中，数据的内容质量、时效性是共享服务的基础，是评估共享服务效果的基础性要素，供给的及时性是影响共享服务质量的前提性因素。Y. W. Lee等[72]给出的数据质量的定义是“反映出数据对特定应用的满足程度”。王志强等[73]结合我国科学数据开发、利用和共享的实际国情，构建了科学数据的质量标准体系，包括数据形式质量、内容质量、效用质量3个一级指标和14个二级指标。从提高共享服务的数据质量、时效性来看，要求科学数据资源来源可靠，即真实、准确，可以不断得到补充和更新。从保证供给的及时性来看，可以通过数据开放平台自身的检索服务、搜索引擎、社交网络等各种传播渠道和途径确保数据用户能够方便快速地定位和获取已开放共享的数据资源，让数据资源快速有效地流通起来。

2.3.2 数据共享服务效果 科学数据只有在广泛共享与应用过程中，才能实现数据拥有者对其获取最大效用的追求。用户满意度是数据共享服务效果评估的一个核心性的因素，用户对开放数据的质量、可用性和价值等有更加直观的感受。黄鑫[67]从科学数据介绍服务、导航服务、获取服务、元数据创建服务、存储服务、引用服务、出版服务、咨询服务和培训服务9个方面探索了质量感知、价值感知和用户满意度三者在科学数据服务领域的相互影响关系，研究结果表明，质量感知对用户满意度的影响非常显著，价值感知对于用户满意度的影响也非常显著。数据资源管理者应密切关注用户数据共享的发展需求，顺势而为，利用自身优势为目标群体提供满足其需求的数据产品与服务，努力提高服务的有用性、易用性，使用户“搜”有所得，“享”有所获。以用户需求驱动，利用各种技术手段，优化数据管理系统与用户的交互设计，完善数据共享服务功能，增加服务的粘度和用户忠诚度，让用户感受到科学数据共享服务功能的实用性和便捷性，提高用户感知满意度，形成有亲和力的用户群。

科学数据共享服务对科技、经济和社会的贡献是评估其服务效果的重要标准，共享主体参与共享活动一个最为明显的动力是参与共享而获得收益，这种收益可以来自三个方面：①对国家各类科技计划和政府决策的支撑及其重大成果的产出有贡献；②对企业的

科技进步和技术创新、教育、科普有支撑作用;③对减少重复投入有作用。

综上,客观准确地对科学数据共享服务效果及其用户满意度进行评价,有利于提升数据资源作为一项学术成果的社会认可度,有助于提高数据工作者的学术地位和影响,最终激发数据工作者出版数据的内在动力。用户数据共享的满意度对形成"人人都是数据使用者和贡献者"的志愿数据共享氛围,促进社会协作,提升科研投资效率,进而对提高科技水平、经济发展、社会效益具有重要推动作用。

3 结论与未来研究展望

3.1 结论

基于对现有相关文献的回顾与梳理,本文得出以下结论:

(1)为了深刻认识和有效揭示科学数据共享研究主题,笔者构建了科学数据共享管理分析框架,该框架描述了发展需求、科学数据共享管理模式与机制、服务共享度与其效果之间的因果关系,强调了科学数据共享管理与服务的合理面、重要度、关键性。

(2)科学数据共享既是数据密集型现代科学研究获取数据的重要途径,也是推进科技进步与创新的必然选择,更是国家发展的战略需求。共享发展需求源于国家战略发展、科学共同体和科研团队三个层面,不同层面的共享需求能够在一定程度上激发、推动科学数据共享活动发展,最终体现了在科学数据开放共享活动中追求实现科学数据价值的共同目标。从国家发展战略需求来看,它是数据资源管理者制定科学数据共享管理整体规划的执行基础,也是建立健全以科学数据资源共享为核心的管理机制的重要前提,更是创建科学数据共享新秩序的根本保证。

(3)科学数据共享具有科学数据管理生命周期的过程性,历经数据生产、组织、出版、传播与利用 5 个步骤,是与科学数据生产、组织、出版、传播、利用直接关联的一种价值创造活动,涉及科学数据的生产者、资助者、组织者、出版者、传播者、管理者、利用者等利益相关主体。对于科学数据共享管理生命周期的价值创造活动而言,有两种力量促进着数据共享价值创造高效完成:科学数据共享的内部动力条件和科学数据共享的外部支撑条件。从数据管理机构可持续发展来看,需要重视管理体制和管理模式的构建,需要强化协调机制,突出共同利益,通过共建共享以集群去消除个别壁垒,以及必须要建立一支服务意识强、高水平的管理技术队伍,这支队伍是科学数据资源取得最大效用价值的充分条件。科学数据共享外部支撑条件强调共享体系的运行效率与保证,体现科学数据资源共享配置强度与能力,也即促进共享活动开展的相关组织机构、制度、信息技术基础设施等外部支撑条件越完善,共享活动运行就越高效。这也意味着促进科学数据共享体系的建设与运行需要综合性措施,既要激发共享主体参与共享的动力,又要完善相关共享组织的制度建设[74]。

(4)科学数据管理共享服务绩效评估体现的是数据管理系统产出的特征,是数据资源服务能力与服务效果的综合。

3.2 未来研究展望

科学数据共享相关研究在一定程度上取得了较好进展,但是基于科学数据共享实践的复杂多样,科学数据共享管理与服务的研究在理论和实践中仍有很大的发展空间。综合来看,国内外现有研究尚存不足,未来研究在加强实证研究、着眼管理机制完善的同时,还需注重服务模式的创新。具体如下:

3.2.1 加强科学数据共享管理的实证研究 一是科学数据价值共创参与者价值贡献度的测定,建立价值贡献评价关键指标,包括使用价值、效用价值等的测定;二是对参与者行为与价值共创绩效关系的实证研究,分析价值共创参与者投入产出关系;三是用户数据共享行为的实证检验,从用户对共享平台服务感知有用性角度来看,数据内容的质量和数量、呈现形式的生动性和互动性等都会影响用户数据共享的行为,它们是如何作用于用户数据共享行为的?这些问题对于科学数据共享服务实践具有重要的现实意义,都有待未来的实证研究去揭示。

3.2.2 完善科学数据共享管理机制 科学数据共享活动涉及科学数据的生产者、资助者、组织者、出版者、传播者、管理者、利用者等各类利益主体。如何吸引不同主体主动参与科学数据共享管理系统的价值共创,如何协调价值共创过程中资源整合的利益与矛盾,如何保证共享管理系统稳定和持续健康发展等都是科学数据管理价值共创面临的核心问题。因此,如何通过机制创新连接促进参与数据共享价值共创主体之间资源整合和互动行为之间的相互影响,促进科学数据服务系统参与者价值共创的动力机制、效益分配机制和协调机制等问题值得进一步关注。

3.2.3 创新科学数据共享服务模式 科学数据的价值实现与现代信息技术的应用密切相关。随着文献资

源数字化向数据化、语义化、关联化和智能化的递阶解构和重构方向发展,数据资源建设的重心将转移到细粒度知识元揭示与黏合服务上来,科学数据共享服务的重心将转移到数据驱动的精细服务和智能评价上来,精细服务在改进用户体验的同时,也在挑战现有的服务利用模式。因此,科学数据共享服务需要有新突破、新发展、新价值,要求我们重新思考科学数据共享服务的核心能力和提升其社会贡献。如何创新科学数据共享服务模式,实现数字资源从内容数据化到数据智能化的价值开发和智慧洞见,值得进一步探讨。

参考文献:

[1] 百度百科[EB/OL].[2019－10－02]. http://baike.baidu.com/item/科学大数据/17648505? fr = aladdin#reference-[2]-17764204-w.

[2] 汪俊. 美国科学数据共享的经验借鉴及其对我国科学基金启示:以 NSF 和 NIH 为例[J]. 中国科学基金,2016(1):69－75.

[3] MULLIGAN A, MABE M. The effect of the Internet on researcher motivations, behavior and attitudes[J]. Journal of documentation, 2011,67(2):290－311.

[4] 司莉,王雨娃. 我国科学数据共享平台数据组织的现状及改进建议——基于国家科技基础条件平台的分析[J]. 图书馆建设,2018(10):52－58.

[5] 邱春艳,黄如花. 近 3 年国际科学数据共享领域新进展[J]. 图书情报工作,2016,60(2):6－14.

[6] 傅小锋,李俊,黎建辉. 国际科学数据的发展与共享[J]. 中国基础科学, 2007(2):30－35.

[7] 美国科学数据中心管理模式调研与分析[EB/OL].[2019－10－06]. https://max.book118.com/html/2018/1001/5333210131001320.shtm.

[8] 促进大数据发展行动纲要[EB/OL].[2019－10－06]. https://baike.so.com/doc/10955176-11483156.html.

[9] 刘垠. 我国首个科学数据管理办法出炉:国家科技计划项目产生的科学数据必须上交[N]. 科技日报. 2018－04－09(1).

[10] 司莉,王雨娃. 我国科学数据共享平台数据组织的现状及改进建议——基于国家科技基础条件平台的分析[J]. 图书馆建设,2018(10):52－58.

[11] 黄鼎成. 科学数据共享的理论基础与共享机制[J]. 中国基础科学,2003(2):22－27.

[12] 盛小平,吴红. 科学数据开放共享活动中不同利益相关者动力分析[J]. 图书情报工作,2019,63(17):40－50.

[13] COUNCIL N, EBRARY I. A question of balance : private rights and the public interest in scientific and tecnhical databases[M]. National Academies,1999.

[14] 陈军,王春聊. 关于科学数据共享机制的思考[J]. 中国基础科学,2003(1):40－43.

[15] CHILVERS A, FEATHER J. The management of digital data: a metadata approach[J]. The electronic library,1998, 16(6):365－372.

[16] 白如江,冷伏海. "大数据"时代科学数据整合研究[J]. 情报理论与实践,2014,37(1):94－99,38.

[17] 孔丽华,刁妍,张晓林. 数据出版的趋势、机制与挑战[J]. 中国科学基金, 2019(3):237－245.

[18] SAYOGO D S,PARDO T A. Exploring the determinants of scientific data sharing: understanding the motivation to publish research data [J]. Government information quarterly,2013,30(S1): 19－31.

[19] 黄国彬,王舒,屈亚杰. 科学数据出版模式比较研究[J]. 大学图书馆学报,2018,36(1):34－40.

[20] 邢文明,刘婷. 增强出版驱动的科学数据出版: 动因、模式及路径[J]. 中国科技期刊研究,2019,30(8):853－860.

[21] 宋立荣,王健. 基于数据服务视角的科学数据出版模式思考[J]. 情报杂志,2019,38(5) : 154－160.

[22] NEVZOROVA O,ZHILTSOV N,ZAIKIN D, et al. Bringing math to LOD: a semantic publishing platform prototype for scientific collections in mathematics[M]// Advanced information systems engineering. Berlin: Springer,2013:379－394.

[23] 邱春艳. 科学数据出版的关键问题研究进展[J]. 图书与情报, 2017(4):54－60.

[24] 涂志芳. 科学数据出版的基础问题综述与关键问题识别[J]. 图书馆,2018(6):86－92,100.

[25] 涂志芳,刘兹恒. 国外数据知识库模式的数据出版质量控制实践研究[J]. 图书馆建设,2018(3):5－13.

[26] 吴立宗,涂勇,王亮绪,等. 浅谈科学数据出版中的数字对象唯一标识符[J]. 中国科技资源导刊,2010,42(5):22－29.

[27] 徐丽芳,丛挺. 数据密集、语义、可视化与互动出版: 全球科技出版发展趋势研究[J]. 出版科学,2012,20(4):73－80.

[28] 徐雷,潘珺. 科学出版物语义数据及其应用研究[J]. 中国科技期刊研究,2018,29(7):704－710.

[29] LAWRENCE B, JONES C,MATTHEWS B, et al. Citation and peer review of data: moving towards formal data publication[J]. International journal of digital curation,2011,6(2):4－37.

[30] SMITH V S. Data publication: towards a database of everything [J]. BMC research notes,2009,2(1):113.

[31] 王雪,马胜利,余曾溧,等. 科学数据的引用行为及其影响力研究[J]. 情报学报,2016,35(11):1132－1139.

[32] 旺丹丹. 科学数据规范引用关键问题探析[J]. 图书情报工作, 2015,59(8) : 42－47.

[33] 史雅莉. 科学数据引用标准实施的关键问题探析. 现代情报, 2019,39(4):34－41.

[34] STARR J, CASTRO E, CROSAS M, et al. Achieving human and machine accessibility of cited data in scholarly publications[J]. PeerJ computer science, 2015, 1(1):e1.

[35] 秦顺,汪全莉,邢文明. 欧美科学数据开放存取出版平台服务调研及启示. 图书情报工作,2019,63(17):129－136.

[36] FORCE M M, ROBINSON N J. Encouraging data citation and discovery with the data citation index[J]. Journal of computer-aided molecular design,2014,28(10):1043－1048.

[37] PARK J H, CHOI S H, KIM B K. A study on utilization of Korea science citation database(KSCD) based on data mining techniques[J]. Journal of information management, 2012, 43(4):191-210.
[38] DATACITE. Citation formatter[EB/OL].[2019-10-01]. https://datacite.org/citation.html.
[39] 丁楠, 杨柳, 丁莹, 等. 我国社会学期刊论文数据引用行为研究[J]. 图书与情报,2014(6):88-93.
[40] 丁楠, 丁莹, 杨柳, 等. 我国图书情报领域数据引用行为分析[J]. 中国图书馆学报,2014, 40(6):107-116.
[41] 邱均平, 何文静. 科学数据共享与引用行为的相互作用关系研究[J]. 情报理论与实践,2015, 38(10):1-5.
[42] 史雅莉,司莉. 我国科研人员数据引用行为特征分析. 情报理论与实践,2019,42(6):36-41.
[43] 王卫华, 胡良霖, 沈志宏. 科学数据引用规范的研制[J]. 中国科技资源导刊,2013, 45(1):36-40.
[44] 王丹丹. 科学数据规范引用关键问题探析[J]. 图书情报工作,2015, 59(8):42-47, 53.
[45] 张静蓓, 仁树怀. 科研数据出版模式、流程及引用策略研究[J]. 图书情报工作,2015, 59(9):21-27.
[46] 黄如花, 李楠. 国外科学数据引用规范调查分析与启示[J]. 图书馆学研究,2016(10):2-9.
[47] 彭洁, 贺德方, 张英杰. 数字出版环境中科学数据引用的实现路径及策略调查分析[J]. 出版发行研究,2014(4):57-61.
[48] 盛小平,武彤. 国内外科学数据开放共享研究综述[J]. 图书情报工作,2019,63(17):6-11.
[49] YATI Y, JAIN R C. Integration of data mining and operation research[J]. Ultra scientist of physical sciences,2009,21(1):311-313.
[50] BERKO Q. Use of ontologies in the processes of semantic data integration[J]. Bulletin of the National University Leiviska Poly Technika,2009,6(5):18-27.
[51] WANG C, LU J, ZHANG G. Web ontology data matching for integration: method and frame work[J]. International journal of Web information systems,2009,5(2):220-238.
[52] 徐枫,李集明. 科学数据共享技术平台构想[J]. 中国基础科学,2003(1):50-55.
[53] Australian code for the responsible conduct of research[EB/OL]. [2019-11-20]. https://research-integrity.uq.edu.au/files/3792/r39_australian_code_responsible_conduct_research_150811.pdf.
[54] Dissemination and sharing of research results[EB/OL]. [2019-11-20]. http://www.nsf.gov/bfa/dias/policy/dmp.jsp.
[55] ESRC research data policy[EB/OL].[2019-11-20]. http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx.
[56] 路鹏,苗良田,莫纪宏,等. 科学数据共享领域的政策规范和法律规范[J]. 国际地震动态,2008(4):35-41.
[57] 张晓青,盛小平. 国外科学数据开放共享政策述评[J]. 图书馆论坛,2018(8):147-154.
[58] 张晓林. 颠覆数字图书馆的大趋势[J]. 中国图书馆学报,2011(9):4-12.
[59] 科技部. 国家科技基础条件平台运行服务绩效考核指标[EB/OL]. [2019-10-13]. http://www.most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj2011/201108/W02.
[60] 基础科学数据网. 共享服务评价指标体系[EB/OL].[2019-10-13]http://www.doc88.com/p-7784333880914.html.
[61] 司莉, 李月婷, 邢文明,等. 我国科学数据共享平台绩效评估实证研究[J]. 图书馆理论与实践, 2014(09):30-35.
[62] 董诚, 赵伟, 涂勇. 我国科学数据机构共享绩效评估研究[J]. 中国科技论坛, 2007(08):75-79.
[63] RUST R T, INMAN J J, JIA J, et al. What you don't know about customer perceive quality: the role of customer expetation distributions[J]. Marketing science,1999,18(1):77-92.
[64] OLIVER R L. Varieties of value in the consumption satisfaction response[J]. Advances in consumer research volume,1996,23(1): 143-147.
[65] 韦草原,王健,张贵兰,等. 基于扎根理论的科学数据用户感知价值概念模型研究[J]. 情报杂志,2018,37(5):182-188.
[66] 严炜炜,张敏. 科研协同中的数据共享与利用行为模式分析[J]. 情报理论与实践,2018,41(1):55-60.
[67] 黄鑫. 基于服务内容的科学数据服务用户满意度研究[D]. 武汉:武汉大学,2017.
[68] 诸云强,朱琦,冯卓,等. 科学大数据开放共享机制研究及其对环境信息共享的启示[J]. 中国环境管理,2015(6):38-45.
[69] 汪俊. 美国科学数据共享的经验借鉴及其对我国科学基金启示:以NSF和NIH为例[J]. 中国科学基金,2016(1):69-75.
[70] 吴丹,陈晶. 我国医学从业者科学数据共享行为调查研究[J]. 图书情报工作,2015,59(18):30-39.
[71] 严炜炜,张敏. 科研协同中的数据共享与利用行为模式分析[J]. 情报理论与实践,2018,41(1):55-60.
[72] LEE Y W, STRONG D M, Knowing-why about date processes and data quality[J]. Journal of management information system, 2003, 20(3):13-39.
[73] 王志强,杨青海. 科学数据质量及其标准化研究[J]. 标准科学,2019(3):25-30.
[74] 郭宝,卓翔芝. 共享链视角下的地区共享资源配置研究[J]. 唐山师范学院学报,2019,41(4):94-100.

作者贡献说明:

毕达天:提出研究思路,撰写论文及定稿;
曹冉:文献整理,撰写及修改论文;
杜小民:文献搜集及整理。

毕达天，曹冉，杜小民. 科学数据共享研究现状与展望[J]. 图书情报工作,2019,63(24):69 - 77.

A Summary of Researches on Research Data Sharing and Development Trends at Home and Abroad

Bi Datian[1] Cao Ran[1] Du Xiaomin[2]

[1] School of Management, Jilin University, Changchun 130022

[2] Yingkou Institute of Technology, Yingkou 115014

Abstract: [**Purpose/significance**] Scientific data sharing has become a strategic demand for national development. Analyzing the research status and development trends of scientific data sharing provides reference for promoting relevant theoretical research and practice in China. [**Method/process**] Through combing and inducing domestic and foreign literatures, three major themes related to scientific data sharing research were extracted, namely "scientific data sharing development needs", "scientific data management sharing modes and mechanisms" and " scientific data service sharing and service effect". We also built the data sharing analysis framework, which reflects the causal relationship among the three themes, and emphasizes the rationality, importance and criticality of scientific data sharing management and service. [**Result/conclusion**] The existing research at home and abroad is still insufficient. In the future, while strengthening empirical research and focusing on the improvement of management mechanism, we should also pay attention to the innovation of shared service mode.

Keywords: scientific data sharing data management data service